

Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise

Mahmoud Keshavarzi^a, Mikolaj Kegler^a, Shabnam Kadir^b, Tobias Reichenbach^{a,*}

^a Department of Bioengineering and Centre for Neurotechnology, Imperial College London, South Kensington Campus, SW7 2AZ, London, UK

^b School of Engineering and Computer Science, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK

ARTICLE INFO

Keywords:

Neural entrainment
Delta and theta frequency bands
Transcranial alternating current stimulation
Speech envelope
Speech comprehension
Speech-shaped-noise
Normal hearing

ABSTRACT

Auditory cortical activity entrains to speech rhythms and has been proposed as a mechanism for online speech processing. In particular, neural activity in the theta frequency band (4–8 Hz) tracks the onset of syllables which may aid the parsing of a speech stream. Similarly, cortical activity in the delta band (1–4 Hz) entrains to the onset of words in natural speech and has been found to encode both syntactic as well as semantic information. Such neural entrainment to speech rhythms is not merely an epiphenomenon of other neural processes, but plays a functional role in speech processing: modulating the neural entrainment through transcranial alternating current stimulation influences the speech-related neural activity and modulates the comprehension of degraded speech. However, the distinct functional contributions of the delta- and of the theta-band entrainment to the modulation of speech comprehension have not yet been investigated. Here we use transcranial alternating current stimulation with waveforms derived from the speech envelope and filtered in the delta and theta frequency bands to alter cortical entrainment in both bands separately. We find that transcranial alternating current stimulation in the theta band but not in the delta band impacts speech comprehension. Moreover, we find that transcranial alternating current stimulation with the theta-band portion of the speech envelope can improve speech-in-noise comprehension beyond sham stimulation. Our results show a distinct contribution of the theta- but not of the delta-band stimulation to the modulation of speech comprehension. In addition, our findings open up a potential avenue of enhancing the comprehension of speech in noise.

1. Introduction

Speech is a complex signal that unfolds over several temporal scales, from phonemes to syllables, words, and phrases. The neural activity in the auditory cortex entrains to the amplitude modulations in speech, as well as to more specific speech structures such as phonemes, the onset of words, and to higher-level linguistic information such as surprisal of word sequences and syntactic structure (Lakatos et al., 2005; Ding and Simon, 2012; Giraud and Poeppel, 2012; Di Liberto et al., 2015; Ding et al., 2016; Brodbeck et al., 2018; Broderick et al., 2018; Weissbart et al., 2019). This cortical entrainment has recently been shown to play a functional role in speech processing. In particular, transcranial alternating current stimulation, paired to rhythmic speech, modulated neural responses that correlated with behaviour when speech was intelligible, but not when it was unintelligible (Zoefel et al., 2018). Moreover, transcranial alternating current stimulation with the speech envelope was found to modulate the comprehension of degraded speech (Riecke

et al., 2018; Wilsch et al., 2018; Kadir et al., 2020). However, it remains unclear which more specific aspects of the cortical speech entrainment underlie the modulation of speech comprehension.

Two main frequency bands dominate the neural speech entrainment. First, cortical activity in the theta frequency band (4–8 Hz) tracks the onset of syllables which may aid the parsing of a speech stream (Ding and Simon, 2014; Di Liberto et al., 2015). A computational model of theta oscillations coupled to gamma oscillations showed indeed that the entrainment of theta activity to a speech signal can act as an efficient parser of syllables, and that the connected gamma network can encode speech efficiently (Hyafil et al., 2015). Second, cortical activity in the delta band (1–4 Hz) entrains to the onset of words in natural speech and has been found to encode both syntactic as well as semantic information (Ding et al., 2016; Broderick et al., 2018; Weissbart et al., 2019).

Much effort has been devoted to tease apart the roles of cortical entrainment in the delta and in the theta band for speech processing (Ding and Simon, 2014; Kösem and Van Wassenhove, 2017). In

* Corresponding author.

E-mail address: reichenbach@imperial.ac.uk (T. Reichenbach).

particular, an MEG investigation into speech with a degraded spectro-temporal fine structure found that the neural entrainment in the delta, but not in the theta, band correlated with speech comprehension (Ding et al., 2014). We have recently employed an experimental paradigm with native and foreign speech in different levels of background noise that allowed us to tease apart the effects of lower-level acoustics and higher-level comprehension, demonstrating that speech acoustics related mostly to theta-band activity and comprehension to delta-band entrainment (Etard and Reichenbach, 2019). These findings agree with a role of the theta band in tracking lower-level acoustical structures such as syllable onsets, and a role of the delta band in entraining to higher-level linguistic features such as semantic and syntactical structures (Ding et al., 2016; Brodbeck et al., 2018; Broderick et al., 2018; Weissbart et al., 2019). However, the distinct roles of both frequency bands to the modulation of speech comprehension through neurostimulation have not yet been investigated.

Here we combined transcranial alternating current stimulation with a behavioural task of speech-in-noise comprehension to tease apart the individual contributions of the delta- and theta band entrainment to speech processing. In particular, we presented young adult participants without hearing impairment with semantically unpredictable sentences that were embedded in speech-shaped noise, such that subjects understood roughly 50% of the words correctly (Fig. 1). Simultaneously to the sound presentation, we stimulated both their left and right auditory cortex symmetrically through small alternating electric currents that were applied through scalp electrodes (transcranial alternating current stimulation or tACS). The current signal was obtained from the envelope

of the simultaneously-presented speech signal. To distinguish between the roles of delta- and theta-band entrainment, we filtered the speech envelope in both frequency bands. We hypothesized that the theta-band and delta-band stimulation would modulate speech comprehension in different ways, since the theta-band stimulation would act on the lower-level acoustic processing while the delta-band stimulation would relate to higher-level linguistic information.

Previous investigations of the modulation of speech comprehension through neurostimulation have partly employed speech that was artificially produced to exhibit a rhythm at a particular frequency (Riecke et al., 2018; Zoefel et al., 2018). These studies then employed an alternating current at the same frequency, and investigated how phase differences between the current and the speech affected comprehension. Alternatively, previous studies used naturalistic speech, the envelope of which had a broad spectrum, and then considered a current waveform that mimicked the speech envelope, but was shifted by different temporal delays (Riecke et al., 2018; Zoefel et al., 2018).

Because we sought to investigate the influence of the neurostimulation in the delta and theta band on speech comprehension, we presented subjects with naturalistic sentences that had significant amplitude modulation in both the delta and the theta frequency range (Fig. 1). We concurrently applied transcranial alternating current stimulation with a waveform that corresponded either to the delta-band portion of the speech envelope or to the theta-band portion of the speech envelope. However, particular care needs to be taken in the analysis of the resulting effects on speech comprehension to avoid analytical bias and false positive results (Asamoah et al., 2019). To avoid

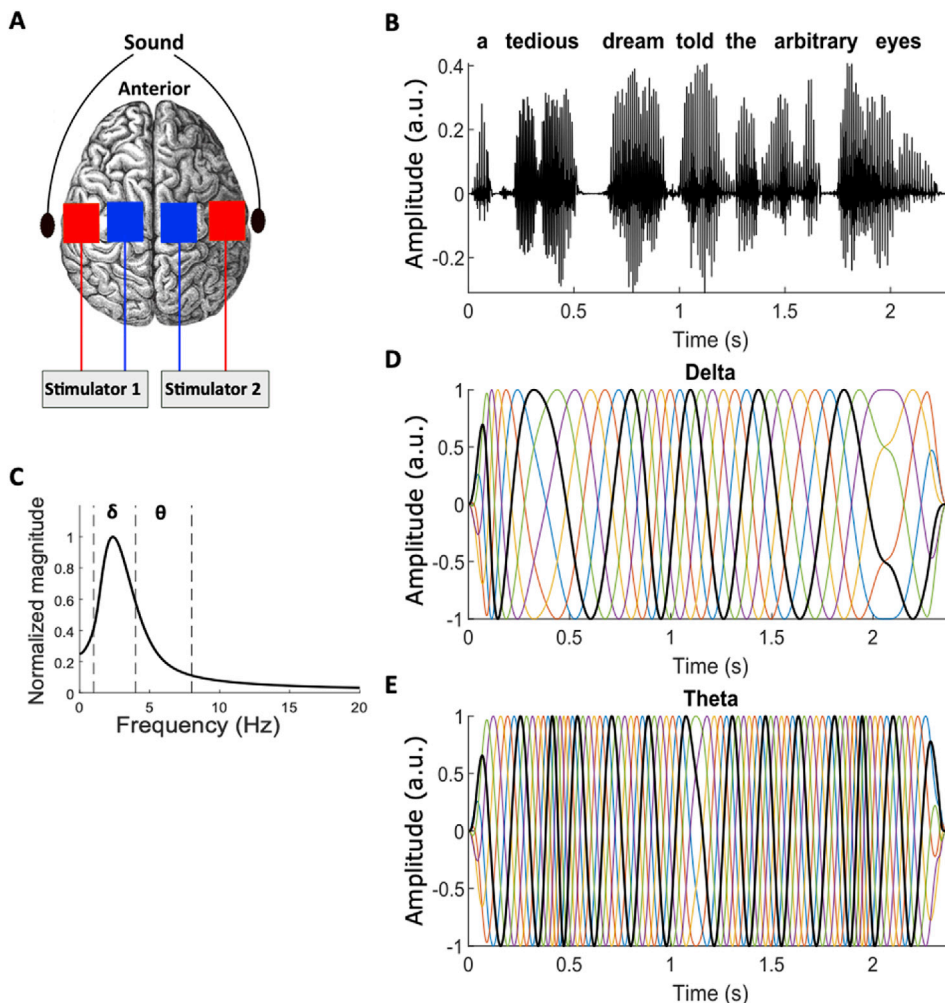


Fig. 1. The experimental design. (A), Participants listened to a sentence embedded in speech-shaped noise. Transcranial alternating electrical current was simultaneously applied symmetrically to both hemispheres through electrodes located over the temporal areas (T7, T8, red) as well as adjacently left and right of the vertex (Cz, blue). (B), Each sentence lasted around 2 s. (C), The spectrum of the envelope of the sentences (computed from averaging over 1000 sentences) was dominated by the delta frequency band, but also contained significant contributions from the theta band. (D, E), We employed current waveforms that followed the speech envelope but were filtered in the delta band (D) or the theta band (E). The resulting waveforms were then shifted by different phases (different colours, black corresponds to no phase shift). The waveforms were further processed so their maxima all had the same value, and such that the values of the minima were all equal as well, except those near the beginning or end of the sentence.

such analytical bias, we used various phase shifts instead of temporal shifts of the current signal. In contrast to temporal changes, phase shifts lead to circular changes of the signal that allowed us to employ powerful methods from Fourier analysis to determine the modulation of speech comprehension.

2. Materials and methods

2.1. Participants

Eighteen native English speakers took part in the experiment (9 females, 8 males, aged between 18 and 29 years, mean age 23 years, standard deviation 3.3 years). All reported normal hearing, had no history of mental health problems or neurological disorders, and were right-handed according to their own assessment. All participants gave informed consent. The experiment was approved by the Imperial College Research Ethics Committee. One female participant did not complete the study due to problems with the electrode attachment.

2.2. Data and code availability

Data and code will be made available on request.

2.3. Hardware setup

A PC with a Windows 7 operating system was used to generate the acoustic stimuli and the current waveforms digitally. Both signals were synchronized on the PC, and were then converted to analogue signals using a USB-6212 BNC device that kept the temporal alignment between the two signals (National Instruments, U.S.A.). The current waveforms were fed to a splitter connected to two neurostimulation devices (DC-Stimulator Plus, NeuroConn, Germany). The acoustic stimuli were passed through a soundcard (Fireface 802, RME, Germany) connected to earphones (ER-2, Etymotic Research, U.S.A.). The temporal alignment of the resulting sound signal to the current waveform was verified by measuring both signals simultaneously, which showed that the timing of both signals differed by less than 1 ms.

2.4. Acoustic stimuli

The acoustic stimuli used in the experiment were single sentences presented in speech-shaped noise. The sentences were semantically unpredictable and were generated using Python's Natural Language Toolkit (Bird et al., 2009; Beysolow, 2018). Each sentence (e.g. "The current months solve the important trial.") consisted of seven words, including five key words used to evaluate the participant's level of comprehension. The sentences were converted to an audio stimulus using the TextAloud software with a male voice and with the sampling rate of 44,100 Hz. The speech signal was presented at an intensity of 65 dB SPL which provided a comfortable sound level.

The speech-shaped noise was generated by determining the average Fourier transform of the different sentences. The phases of the spectral components were then randomized while the magnitude was kept. The noise was then obtained by the inverse Fourier transform of the resulting randomized signal.

2.5. Neurostimulation waveforms

We presented subjects with speech signals and concurrently applied transcranial alternating current stimulation. For the latter we employed 15 different waveforms. One waveform was designed to provide a sham stimulus. This current started at the beginning of the speech signal but lasted only 500 ms. Smooth onsets and offsets were produced through ramps of a duration of 100 ms. This sham stimulation was used to mimic the current delivery, in particular the attachment of the scalp electrodes. It could in principle also control for a brief skin sensation resulting from

the current, although, as described below, we adjusted the current magnitude such that subjects did not experience a skin sensation.

The other 14 waveforms were all based on the speech envelope. The latter was computed as the absolute value of the analytical signal of the speech. The speech envelope was then band-pass filtered into the delta frequency band (zero phase IIR filter, low cutoff (−3 dB) 1 Hz, high cutoff (−3 dB) 4 Hz, order 6). The envelope was also band-pass filtered into the theta frequency band (zero phase IIR filter, low cutoff (−3 dB) 4 Hz, high cutoff (−3 dB) 8 Hz, order 6). The band-pass filters implied that both waveforms had a mean of 0.

To enhance the influence of the current signal on the neural entrainment, the waveforms were then processed to boost all maxima and minima in the waveform to the maximal (minimal) value that was encountered in the signal. This was done by computing the analytical (complex) signal through the Hilbert transform, by subsequently setting the amplitude to unity, and by then taking the real part of the obtained function.

The waveforms in both the delta and theta frequency band was then shifted by the six phases 0°, 60°, 120°, 180°, 240° and 300°. A shift by a phase ϕ was implemented by first computing the analytical signal of the band-pass filtered envelope, followed by multiplication by $e^{i\phi}$ (where ϕ has been converted to radians) and by taking the real part of the obtained signal. Because $e^{i(\phi+2\pi)} = e^{i\phi}$, this procedure ensured the circularity of the phase shifts, despite the broad frequency range of the speech envelope. In particular, a shift by a phase of $\phi + 360^\circ$ (where ϕ is measured in degrees again) yielded the same signal as a shift by phase ϕ .

The six phase shifts of both the delta- and the theta-band envelope yielded twelve waveforms. We furthermore employed a delta-band and a theta-band envelope that were obtained from an unrelated sentence, yielding two more current waveforms.

2.6. Experimental set up and procedure

The participants were seated in a soundproof room. The sound was presented diotically through earphones (ER-2, Etymotic Research, U.S.A.). Two rubber electrodes were placed adjacently left and right of the location Cz of the subject's head, and the remaining two at the locations T7 and T8 of the International 10-10 system (Fig. 1A). One electrode near Cz and the one at T7 were connected to one neurostimulation device, and the remaining electrodes to the second device. Based on simulations of electrical field distribution in a standard human head model and previous experimental studies, such a configuration of electrodes induces strong modulation of the auditory cortices (Herrmann et al., 2013; Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018). The electrodes at the temporal areas served as the anodes and the electrodes at Cz as the cathodes. The electrodes were covered by 35 cm² sponge pads moistened by a 0.9% saline solution (about 5 ml per side). After placing them on the participant's head, the impedance between electrodes of each device was set to below 10 k Ω .

To measure the maximum magnitude of the stimulation current to be used for a participant, a pure sinusoidal signal at a frequency of 3 Hz and with a duration of 5 s was presented to the subject. The signal amplitude was increased from 0.1 mA to a maximum of 1.5 mA in step sizes of 0.1 mA. To minimize the transcutaneous effects of tACS, the procedure was stopped when the participant reported a skin sensation, and the amplitude of the previous step was selected as the maximum threshold for the stimulation current for that participant. The maximal currents that we thereby estimated for the different participants were in the range of 0.7–1.3 mA, with a mean of 1.1 mA and a standard deviation of 0.3 mA.

For each participant, we first measured the sentence reception threshold (SRT) of 50%, that is, the signal-to-noise ratio at which speech comprehension was 50%. During the measurement the participants were subjected to sham stimulation at the onset of each sentence. To estimate the SRT, we employed an adaptive procedure (Kollmeier et al., 1988; Kaernbach, 2001). We started with an initial SNR that was randomly

selected between 0 dB and -3 dB. If the subject understood at least three key words in the sentence correctly, the SNR value was decreased by 1 dB for the subsequent sentence. The SNR was increased by 1 dB otherwise. The adaptive procedure was stopped after seven reversals in the SNR or after 17 sentences. The adaptive procedure was carried out four times for each subject. The subject's SRT was computed as the average of the last three SNRs that were employed in each of the different runs of the adaptive procedure, with the exception of those of the first run. The so-established SRT was then used as the SNR for the subsequent measurements.

We then measured subjects' speech comprehension during concurrent transcranial alternating current stimulation with 15 different waveforms. For each waveform we therefore presented each subject with a total of 25 sentences in speech-shaped noise, at the SNR corresponding to the personalized SRT that was measured earlier, and applied the current stimulation simultaneously. After listening to each sentence, the subject repeated what he or she understood. The response was recorded through a microphone and manually graded by the experimenter for the percentage of correctly understood words. A total of 375 sentences was presented in two different testing sessions that took place on two different days. Which of the 15 different waveforms was used for the current stimulation varied randomly from sentence to sentence and was unknown to both the experimenter and the subject (double blind design). After every 50 sentences the subject took a 2-min break.

2.7. Statistical analysis

To investigate the modulation of speech comprehension through both the delta- and the theta-band neurostimulation, we shifted the envelope in each of the two frequency bands by six different phases (0° , 60° , 120° , 180° , 240° and 300°). Each phase shift can modulate the cortical entrainment in the respective frequency band differently: a particular phase shift may, for instance, increase the cortical entrainment whereas another one may diminish it (Riecke et al., 2018; Zoefel et al., 2018). Importantly, although the band-pass filtered envelopes did contain a range of frequencies, the phase shifts were applied in such a way that they were nonetheless cyclical. In particular, a phase change of 360° corresponded to no phase change at all (0°).

If the current stimulation affected speech comprehension, the latter would depend in a cyclical manner on the phase of the current stimulation. In contrast, a finding of no dependence of speech comprehension on the neurostimulation phase would signal that there is no influence of the stimulation, and hence no impact of the neural entrainment on speech processing. We therefore measured the comprehension scores of volunteers and analyzed their dependence on the phase of the current stimulation.

We performed this analysis separately for the current waveforms filtered in the delta and in the theta frequency band. Because we measured the comprehension scores at different phase shifts, the circularity of the phase, and the resulting circularity of the dependence of the speech comprehension on the stimulation phase, meant that the data could be analyzed using the Discrete Fourier Transform. In particular, the data could be written as a discrete sum of cosine functions, each with a particular period that was the either the largest-possible period of 360° or a fraction of 360° . Because we measured speech comprehension at six different phases $\{\phi_k\}_{k=1}^6$, the Discrete Fourier Transform implied that the dependence of the speech comprehension score $CS(\phi_k)$ on the phase ϕ_k of the current stimulation could be written as

$$CS(\phi_k) = \sum_{n=0}^5 a_n e^{in\phi_k}, \quad (1)$$

with the complex Fourier coefficients a_n . Four of these coefficients are related through complex conjugation: $a_4 = a_2^*$ and $a_5 = a_1^*$. Let $\frac{A_1}{2}$ be the magnitude of the complex coefficient a_1 , and $-\Phi_1$ its phase: $a_1 =$

$\frac{A_1}{2} e^{-i\Phi_1}$. The coefficient a_5 follows via complex conjugation. The coefficient a_2 can be expressed analogously through its amplitude and phase as $a_2 = \frac{A_2}{2} e^{-i\Phi_2}$. The coefficient a_4 follows as the complex conjugate. The two coefficients a_0 and a_3 are real: they denote a constant offset respectively a contribution that alternates at $+1$ and -1 . They are therefore entirely defined by their magnitudes A_0 and A_3 , respectively: $a_0 = A_0$ and $a_3 = A_3$. Because the six discrete phase values $\{\phi_k\}_{k=1}^6$ at which we have assessed speech comprehension lead to $e^{i6\phi} = 1$, we have $e^{in\phi} = e^{i(n-6)\phi}$ and can therefore write equation (1) as

$$CS(\phi) = A_0 + A_1 \cos(\phi - \Phi_1) + A_2 \cos(2\phi - \Phi_2) + A_3 \cos(3\phi). \quad (2)$$

The model parameters A_1 , A_2 and A_3 hereby denote the amplitude of the variation at the periods 360° , 180° and 120° , respectively. The phases Φ_1 and Φ_2 are the phase shifts at the two longer periods. Because the shortest period corresponds to the Nyquist frequency, it does not allow the inference of a phase shift. A_0 denotes a constant offset. The resulting number of parameters is six, matching the number of phase shifts at which comprehension scores are measured.

We determined the offset A_0 from the mean comprehension score. The modulation amplitudes A_1 , A_2 and A_3 as well as the phase shifts Φ_1 and Φ_2 were computed through the Discrete Fourier Transform. We then wondered which of the amplitudes would be statistically significant. Significance of either of these amplitudes would mean that there was a significant dependence of speech comprehension on the stimulation phase at the corresponding period. This would therefore show a significant modulation of speech comprehension through the current stimulation.

The significance of the modulation amplitudes was determined in two independent ways. First, we kept the two phase shifts Φ_1 and Φ_2 as well as the constant offset A_0 fixed, and estimated the amplitudes A_1 , A_2 and A_3 from multiple linear regression. We then determined the associated p -values and corrected for multiple comparisons through the FDR correction.

Second, we employed a permutation-based method to test the significance of the modulation amplitudes A_1 , A_2 and A_3 . We therefore computed null models for these amplitudes. The null models were obtained from random permutations of the speech comprehension scores across the six different phases. The permutations were done separately for each subject. For each set of permutations, the parameters in Equation (1) were then determined from the Discrete Fourier Transform, as for the actual data. We performed this procedure 10,000 times, resulting in 10,000 null models. We therefrom obtained the null distributions of the modulation amplitudes A_1 , A_2 and A_3 . We determined the amplitude threshold such that the probability to have a higher amplitude in a null model was 1.7%. This corresponded to a probability of 5% with a Bonferroni correction for the three comparisons. The Bonferroni correction was employed instead of the FDR correction since the latter requires p -values and could not be employed to obtain an amplitude threshold. The null models further allowed us to compute p -values for the amplitudes. The p -value of a particular amplitude followed as the probability of observing a larger value in a null model.

The phase dependence of speech comprehension may differ from subject to subject. We therefore also analyzed the data when aligning the phase to the 'best phase' per subject, that is, to the phase that yielded the highest speech comprehension score for that particular subject. We then analyzed the speech comprehension scores $CS(\tilde{\phi})$ at the phases $\tilde{\phi}$ that were aligned to the best phase through the Discrete Fourier Transform. Because the alignment with respect to the best phase left us with only five phases, the Discrete Fourier Transform had only five instead of the previous six parameters:

$$CS(\tilde{\phi}) = A_0 + A_1 \cos(\tilde{\phi} - \Phi_1) + A_2 \cos(2\tilde{\phi} - \Phi_2). \quad (3)$$

In particular, a modulation of speech comprehension could arise through a modulation at either the period of 360° or 180° , with the

modulation amplitude of A_1 and A_2 , respectively.

We determined the statistical significance of the two amplitudes A_1 and A_2 as for the case of the non-aligned data described above. In particular, we used two independent methods, multiple linear regression and the permutation-based test.

3. Results

3.1. Relation between time-shifted and phase-shifted waveforms

We first sought to investigate the effect of the phase shifts on the neurostimulation waveforms. Both the neurostimulation signal in the delta frequency band as well as that in the theta frequency band contained a range of frequencies and therefore differed from purely sinusoidal signals (Fig. 1D and E). Because the same phase shift was applied to all frequency components, the phase shift did not change the group delay, which follows as the derivative of the phase with respect to frequency. However, the phase delay is defined as the ratio of the phase to the angular frequency, and was therefore altered by the phase shift, in a manner that varied with the frequency. This effect led to a phase-shifted signal that had a different shape from the original one. Moreover, the phase-shifted signal differed from a time-shifted waveform as well.

However, because both the delta-band portion and the theta-band portion of the speech envelope are comparatively narrow-band signals, phase shifts translated approximately to temporal shifts as long as the latter were not too long. To quantify this correspondence, we computed the cross-correlation between the delta-band signal shifted by different phases and temporal delays with the unshifted version, that is, with the signal with neither a time shift nor a phase shift (Fig. 2A). We found that for latencies around 0 the maximal correlation values were close to +1. As an example, a maximal correlation value of 0.5 (across phases) was observed for delays between -210 ms and 210 ms. If we consider a correlation value of at least 0.5 to denote a reasonable correspondence between two signals, then this shows that time delays between -210 ms and 210 ms could be approximately represented by phase shifts. We carried out the same analysis for the speech envelope filtered in the theta band (Fig. 2B). We obtained maximal correlation (across the different phase shifts) of at least 0.5 for temporal shifts between -150 ms and 150 ms, evidencing that such temporal delays could partly be captured by phase shifts.

The cross-correlation analysis also verified the cyclical nature of the

phase changes. In particular, in the absence of a temporal delay, a signal at a phase change of -180° or of 180° was anti-correlated to the signal without a phase change. The phase change of -180° or of 180° did indeed yield a signal that corresponded to the original one, but with the opposite polarity (Fig. 1D and E). Other phase shifts led to a cross-correlation with the unshifted waveform that changed cyclically from -1 (perfect anti-correlation) for a phase shift of -180° to 0 (no correlation) for a phase shift of -90° , to $+1$ (perfect correlation) for no phase shift (0°), and then back to 0 (no correlation) for a phase shift of 90° and to -1 (perfect anti-correlation) for a phase shift of 180° .

These results confirm that phase shifts and temporal delays are two different ways to manipulate the neurostimulation waveform. Although phase changes relate approximately to temporal delays as long as these are not too long, both manipulations yield in general different results and can therefore have different effects on speech comprehension. In this study we employed phase shifts since this type of manipulation allowed us, due to the cyclical nature of the phase shifts, to use circular statistics for the investigation of the resulting speech comprehension.

3.2. Modulation of speech comprehension through theta- but not delta-band neurostimulation

We measured speech comprehension scores while participants experienced transcranial alternating current stimulation with a waveform that was derived from the speech envelope, but band-pass filtered into either the delta or the theta band. To explore the effect of the two types of current stimulation on speech comprehension, we then employed current waveforms that were shifted by six different phases (0° , 60° , 120° , 180° , 240° and 300°). As set out in the Methods section, due to the cyclical nature of the phase, the dependence of the speech comprehension score on the phase of the current stimulation can be written as a linear combination of sinusoidal variations at periods of 360° , 180° and 120° (Equation (1)). We computed the amplitudes A_1 , A_2 and A_3 of these variations through the Discrete Fourier transform. We then assessed the statistical significance of each modulation amplitude through two independent methods, multiple linear regression as well as a permutation-based test.

For the current stimulation with the speech envelope filtered in the delta band, the multiple linear regression showed that none of the amplitudes were statistically significant ($df = 3$; $A_1 = 0.01$, $t = 2.9$, $A = 0.2$; $A_2 = 0.01$, $t = 1.8$, $p = 0.2$; $A_3 = 0.005$, $t = 1.0$, $A = 0.3$; $R^2 = 0.064$; FDR

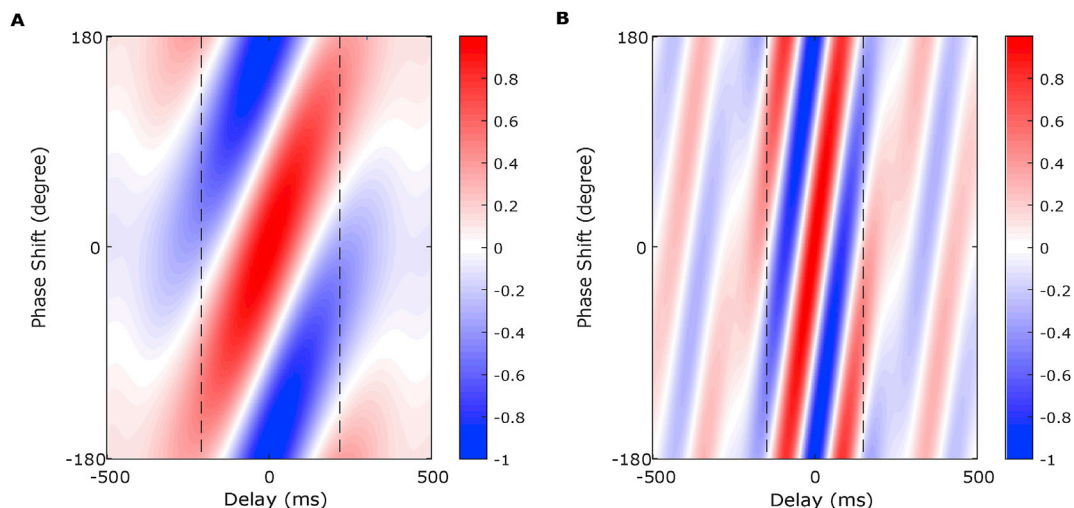


Fig. 2. The relation between phase and time shifts. (A), The correlation of the speech envelope filtered in the delta band, to this signal shifted by different delays and phases. A temporal shift can be compensated by a certain shift in phase. In particular, the maximal correlation (across the different phases) is at least 0.5 between a delay of -210 ms and 210 ms (dashed lines). (B), The correlation of the theta-band filtered speech envelope with a version shifted in phase and time. A time shift can be compensated by a certain phase shift: the maximal correlation (across phase) exceeds 0.5 for delays between -150 ms and 150 ms (dashed lines). For larger temporal shifts there is less correspondence between time and phase shifts.

correction for multiple comparisons, Fig. 3A). This was confirmed by the permutation-based method ($A_1, p = 0.3; A_2, p = 0.1; A_3, p = 0.2$; Fig. 4A–C). There was accordingly no modulation of speech comprehension through the delta-band current stimulation.

For the stimulation in the theta band, however, the multiple linear regression revealed the statistical significance of the modulation amplitude A_1 , although the others were insignificant ($df = 3; A_1 = 0.02, t = 2.9, P = 0.01; A_2 = 0.01, t = 1.5, P = 0.2; A_3 = 0.0002, t = 0.03, P = 0.97; R^2 = 0.097$; FDR correction for multiple comparisons, Fig. 3B). The permutation test corroborated this finding ($A_1, P = 0.01; A_2, P = 0.3; A_3, P = 0.3$; Fig. 4D–F). This showed that the theta-band current stimulation had a significant influence on speech comprehension, namely at the longest period of 360° .

3.3. Consistent phase dependencies across subjects

The above analysis was performed on the population level, and the phase of the neurostimulation was not adjusted per subject. However, prior studies found that the effect of neurostimulation on speech comprehension may depend on the parameters of the current stimulation, such as phase delay or time shift, in a manner that is not consistent across subjects (Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018). We therefore investigated whether we had significant subject-to-subject variation in the dependence of the comprehension scores on the stimulation phase. To this end, we determined for every subject, and separately for the delta and for the theta band, the phase that yielded the highest comprehension score. We referred to this phase as the ‘best phase’ for that subject, and aligned the phase relative to this best phase (Fig. 5A and B). We performed the analysis of the dependence of the comprehension scores on the relative phase through the model given by Eqn (3). This model described the dependence of the speech comprehension scores on the aligned phases through variations at only two periods, 360° and 180° , with the corresponding amplitudes A_1 and A_2 , reflecting that only five phases remain after the alignment to the best phase.

For the stimulation with the delta-band filtered speech envelope, the multiple linear regression revealed no significant modulation of speech comprehension ($df = 2; A_1 = 0.009, t = 1.0, P = 0.3; A_2 = 0.005, t = 1.2, P = 0.3; R^2 = 0.03$); FDR correction for multiple comparisons, Fig. 5A), which was confirmed by the permutation-based test ($A_1, P = 0.3; A_2, P =$

0.2). Likewise, the multiple linear regression showed no significant impact of the theta band stimulation either ($df = 2; A_1 = 0.008, t = 0.35, P = 0.7; A_2 = 0.007, t = 0.80, P = 0.7; R^2 = 0.01$); FDR correction for multiple comparisons, Fig. 5B). This was corroborated by the permutation test ($A_1, P = 0.6; A_2, P = 0.2$). The alignment with respect to the best phase per subject accordingly rendered the previously-obtained modulation with speech comprehension through the theta-band current insignificant.

To investigate the potential inter-subject variability of the phase dependence further, we computed the distribution of the subjects’ best phases (Fig. 5C and D). We found that, for neurostimulation in the delta band, the distribution was not significantly different from a uniform one ($p = 0.4$, Rayleigh test). This accorded with our finding that delta-band stimulation did not have a significant influence on speech comprehension, since the best phase is then distributed randomly. The current stimulation in the theta band, however, showed a distribution of the best phases that differed significantly from uniformity ($p = 0.02$, Rayleigh test). The mean phase was $36^\circ \pm 30^\circ$. This provided additional evidence that the best phase for the theta-band stimulation was consistent across subjects.

3.4. Enhancement of speech comprehension through theta-band neurostimulation

Furthermore, we wondered whether current stimulation could not only modulate but actually enhance the comprehension of speech in noise. We therefore also measured the comprehension scores when subjects experienced a sham stimulus. As an additional control, we stimulated volunteers with a current that followed the envelope of an unrelated sentence, filtered either in the delta or in the theta frequency band. These currents obtained from unrelated sentences should not facilitate speech comprehension, but, if anything, hinder it.

We compared the comprehension scores that we obtained for the delta- and theta-band stimulation at the phase that yielded the highest comprehension across subjects — the phase of 0° in either case — to the different control conditions (Fig. 6). We found that there was statistically significant variation between the different comprehension scores (One-way ANOVA, $df = 4, F = 3.1, p = 0.02, \eta^2 = 0.1$). Post-hoc tests showed that the only two types of neurostimulation that yielded significantly different speech comprehension were the theta-band stimulation and the

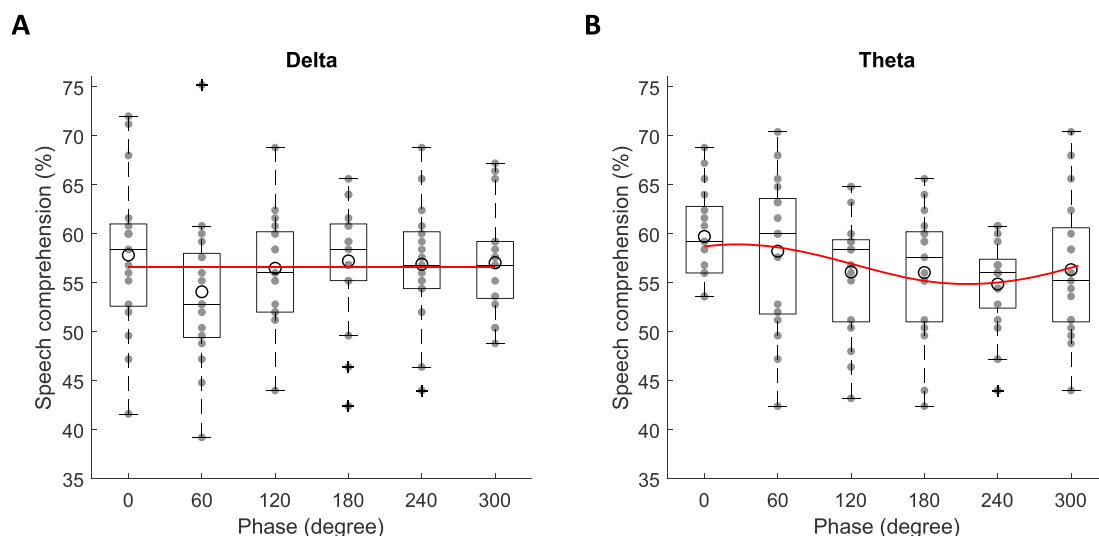


Fig. 3. Modulation of speech comprehension through theta-band but not delta-band current stimulation. Speech comprehension scores at different phases are shown as box plots, the circles indicate the mean values and crossed denote outliers. Results from individual subjects are indicated as grey disks. The red line denotes the fit obtained from the model given by equation (1), but using only those terms that are statistically significant. (A), Speech comprehension during delta-band stimulation is not influenced by the phase of the stimulation. (B), Theta-band stimulation leads to a significant modulation of speech comprehension, at the longest possible period of 360° .

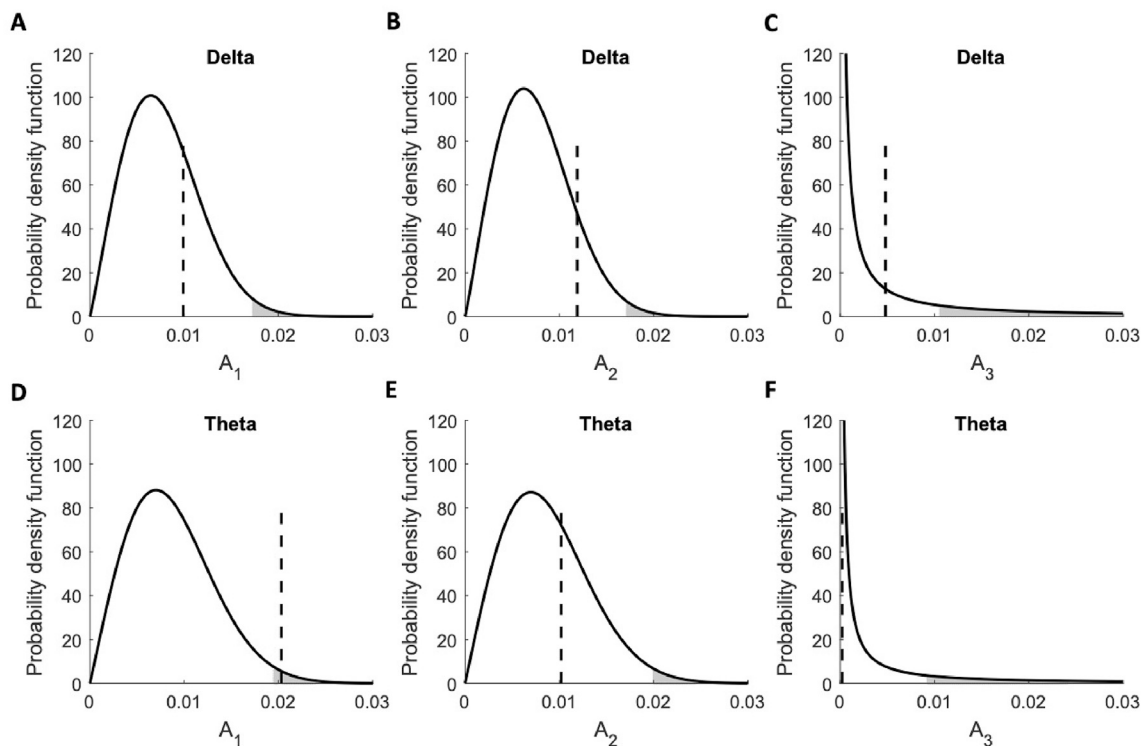


Fig. 4. Significant dependence of speech comprehension on the stimulation phase for theta-band but not delta-band current stimulation. We used permutations of the speech-comprehension scores to compute null models of the modulation amplitudes, and therefrom their probability distributions (black lines). The grey areas show the largest amplitudes that were observed in the null models with a probability of less than 1.7%, which corresponded to a probability of 5% adjusted for the three comparisons with the Bonferroni correction. The modulation amplitudes computed from the actual data are shown as dashed lines (A-C). The dependence of speech comprehension on the stimulation phase for delta-band stimulation is insignificant at all three periods. (D-F) The dependence of speech comprehension on the stimulation phase for the theta-band stimulation is significant for the longest period (A₁ is significant) but not at the two others (A₂ and A₃ are insignificant).

sham stimulation ($p = 0.03$), Tukey-Kramer method (Tukey, 1949; Driscoll, 1996). In particular, transcranial alternating current stimulation with the theta-band filtered speech envelope, and without a phase shift, yielded speech comprehension that was significantly better than the one obtained under sham stimulation. Speech comprehension improved by 6%, which is comparable to the efficacy of some noise-reduction algorithms for hearing aids and suggests that this type of neurostimulation may have practical applications in auditory prosthetics (Chung, 2004; Healy, Eric W., Masood Delfarah, Eric M. Johnson, 2019).

We also wondered if the variances of the speech comprehension scores differed between the various conditions. Although the variance was largest for the delta-band stimulation, we did not find a statistically-significant difference between the five conditions (Bartlett's test, $k = 5$, $\chi^2 = 7.1$, $p = 0.13$).

4. Discussion

We showed that neurostimulation with the theta-band but not the delta-band portion of the speech envelope impacts comprehension. This finding ties in with previous studies that have identified different roles of these two frequency bands for speech processing. In particular, entrainment in the theta band has been shown to relate to acoustic properties of speech, including the clarity of a speech signal in background noise, whereas the delta-band entrainment can inform on higher-level linguistic aspects of speech such as syntactic features, semantics, and thereby comprehension (Ding and Simon, 2014; Di Liberto et al., 2015; Hyafil et al., 2015; Broderick et al., 2018; Weissbart et al., 2019). Our study suggests that the theta-band entrainment plays a functional role, perhaps through aiding the acoustic parsing of speech. Our observed lack of modulation of speech comprehension through delta-band stimulation

may reflect that, although the neural speech tracking in the delta band relates to higher-level linguistic information in speech and to speech comprehension, this relationship originates in only a small portion of the delta-band entrainment (Ding et al., 2016; Broderick et al., 2018; Etard and Reichenbach, 2019). Transcranial alternating current stimulation with the delta-band portion of the speech envelope may not be efficient in modulating this small neural response. Alternatively, the effect may have been too small to observe in the comparatively small number of 17 subjects that we assessed here, or the delta-band speech entrainment may be an epiphenomenon of other neural processes.

Cortical activity entrains to speech rhythms at different temporal lags, in particular at an early latency of 150 ms and a longer latency of 250 ms, suggesting that the timing of the neurostimulation signal with respect to the sound may affect how comprehension is modulated (Horton et al., 2013; Ding and Simon, 2014). Previous studies on the effects of neurostimulation on speech processing have partly investigated different temporal lags between the speech signal and the transcranial alternating current, and found best lags that were distributed broadly among participants between -400 and 400 ms (Riecke et al., 2018; Wilsch et al., 2018). While our approach employed no temporal delay between the envelope-based current and the speech, our analysis showed that the phase shifts that we used partly corresponded to time lags of about 200 ms in magnitude, such that our approach effectively captures a significant range of temporal delays.

We found evidence of a consistent phase, across volunteers, at which the theta-band current stimulation modulated speech comprehension. Moreover, when considering a subject-specific phase alignment, we no longer obtained a significant effect of phase on speech comprehension. This may indicate that the alignment of the phase according to the best phase per subject increased the noise in the data, which may in turn follow from uncertainty in determining the best phase for each

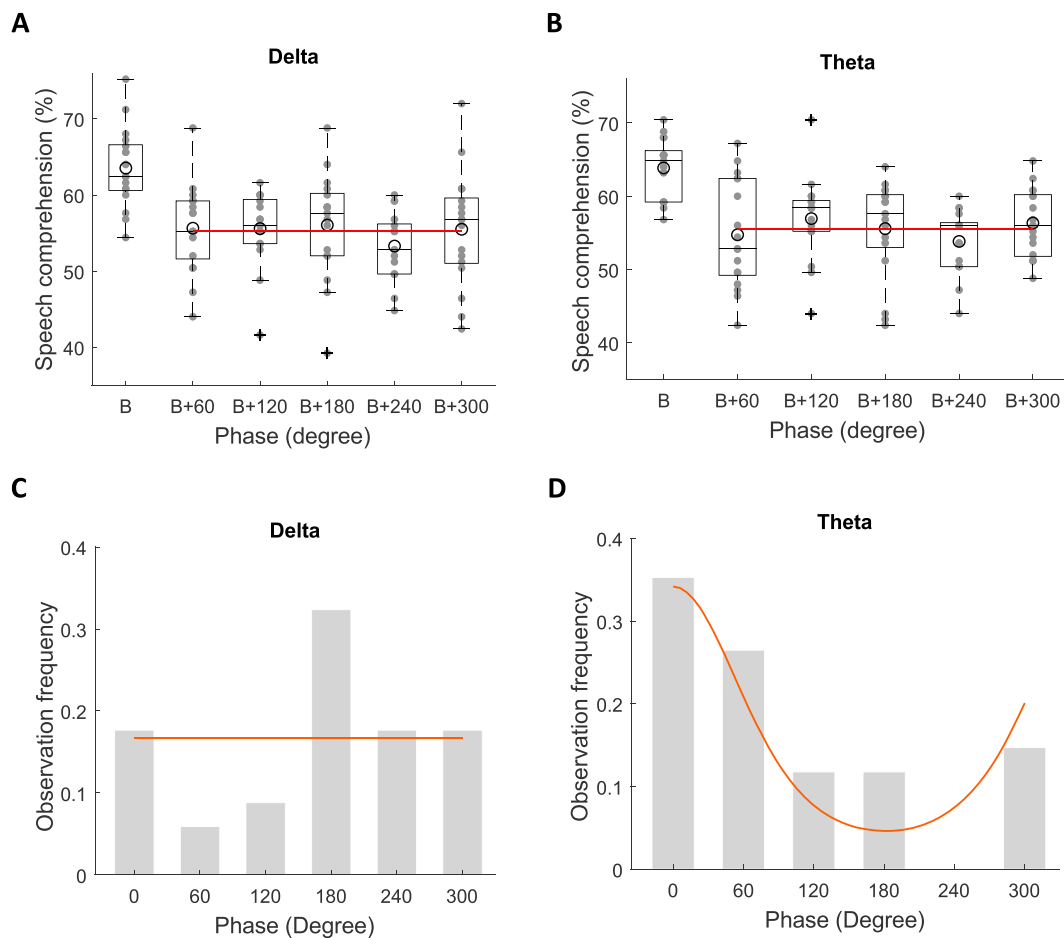


Fig. 5. Consistent phase dependency across subjects. (A, B), Results on the population level are shown through box plots, with crosses denoting outliers. Open circles denote the mean values, and grey disks indicate the results from individual subjects. When adjusting the phase relative to the best phase ('B') per volunteer, neither the delta-band stimulation (A) nor the theta-band entrainment (B) lead to significant effect of phase on speech comprehension. The red line denotes the fit obtained from the significant parts of the model given in Equation (1). (C), For delta-band stimulation the distribution of the best phases (grey bars) is not significantly different from a uniform distribution (red line). (D), Theta-band stimulation leads to a distribution (grey bars) that differs significantly from uniformity. The best phases occur around the mean phase of $36^{\circ} \pm 30^{\circ}$. The distribution can be approximated well by a von Mises distribution (red line).

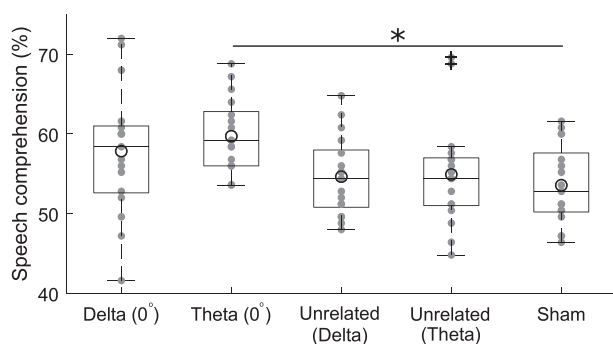


Fig. 6. Enhancement of speech comprehension through current stimulation. We compared current stimulation with the best phase of the delta- and theta-band waveforms (0° for both), stimulation with the envelope of an unrelated sentence filtered either in the delta or in the theta frequency band, as well as sham stimulation. Theta stimulation without phase shift leads to significantly better comprehension scores than sham stimulation. Box plots denote results on the population level, with open circles showing the population mean and crosses indicating outliers. Grey disks show the results from individual subjects. The asterisk indicates a statistically-significant difference.

individual. However, our finding of a consistent influence of phase on speech comprehension across the subjects differed from previous studies that found broad variability in how certain temporal lags or phase shifts modulated speech comprehension (Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018). These studies employed either the broad-band speech envelope, mostly between 1 and 15 Hz, or speech that was artificially altered to follow a single rhythm, which may have increased the variability across participants.

Because the theta-band entrainment plays a functional role in speech comprehension, we expected that current stimulation with an unrelated envelope would worsen speech comprehension compared to a sham stimulus. However, we found that neither stimulation with an unrelated delta band envelope nor with an unrelated theta-band envelope rendered significantly lower comprehension scores. This may indicate that, perhaps due to the relatively high background noise, the theta-band entrainment in the absence of current stimulation was already rather low and did not decrease significantly further upon stimulation with an unrelated envelope.

5. Conclusion

In summary, our results show that the modulation of speech comprehension through transcranial alternating current stimulation stems from the theta but not from the delta band. We have further

demonstrated that the theta-band stimulation modulates speech comprehension in a manner that is consistent across subjects. In particular, there exists an optimal phase shift across subjects at which speech comprehension is aided. Importantly for potential practical applications, our results evidence that current stimulation within the theta frequency band can enhance speech comprehension with respect to sham stimulation, a result that had not been possible with the use of broad-band current stimulation (Riecke et al., 2018; Wilsch et al., 2018).

Declaration of competing interests

The authors declare no competing financial interests.

Acknowledgement

We thank Octave Etard, Pierre Guillemot and Enrico Varano for help with the experimental setup. This research was supported by EPSRC grant EP/R032602/1 to T.R. and by the Royal British Legion Centre for Blast Injury Studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2020.116557>.

References

- Asamoah, B., Khatoun, A., Mc Laughlin, M., 2019. Analytical bias accounts for some of the reported effects of tACS on auditory perception. *Brain Stim* 12, e165.
- Beysolow II, T., 2018. Applied Natural Language Processing with Python.
- Bird, S., Klein, E., Loper, E., 2009. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.
- Brodbeck, C., Hong, L.E., Simon, J.Z., 2018. Rapid transformation from auditory to linguistic representations of continuous speech. *Curr. Biol.* 28, 3976.
- Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.* 28, 803.
- Chung, K., 2004. Challenges and recent developments in hearing aids: Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends Amplif.* 8, 83–124.
- Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25, 2457–2465.
- Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroim* 88, 41–46.
- Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158.
- Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859.
- Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8, 311.
- Driscoll, W.C., 1996. Robustness of the ANOVA and Tukey-Kramer statistical tests. *Comput. Ind. Eng.* 31, 265–268.
- Etard, O., Reichenbach, T., 2019. Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *J. Neurosci.* 39, 5750–5759.
- Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517.
- Healy, Eric W., Delfarah, Masood, Johnson, Eric M., Wang, D., 2019. A deep learning algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker and reverberation. *J. Acoust. Soc. Am.* 145, 1378–1388.
- Herrmann, C.S., Rach, S., Neuling, T., Strüber, D., 2013. Transcranial alternating current stimulation: a review of the underlying mechanisms and modulation of cognitive processes. *Front. Hum. Neurosci.* 7, 279.
- Horton, C., D'Zmura, M., Srinivasan, R., 2013. Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* 109, 3082–3093.
- Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., Giraud, A.-L., 2015. Speech encoding by coupled cortical theta and gamma oscillations. *Elife* 4, e06213.
- Kadir, S., Kaza, C., Weissbart, H., Reichenbach, T., 2020. Modulation of speech-in-noise comprehension through transcranial current stimulation with the phase-shifted speech envelope. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 23.
- Kaernbach, C., 2001. Adaptive threshold estimation with unforced-choice tasks. *Percept. Psychophys.* 63, 1377.
- Kollmeier, B., Gilkey, R.H., Sieben, U.K., 1988. Adaptive staircase techniques in psychoacoustics: a comparison of human data and a mathematical model. *J. Acoust. Soc. Am.* 83, 1852.
- Kösem, A., Van Wassenhove, V., 2017. Distinct contributions of low-and high-frequency neural oscillations to speech comprehension. *Lang Cogn Neurosci* 32, 536–544.
- Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911.
- Riecke, L., Formisano, E., Sorger, B., Başkent, D., Gaudrain, E., 2018. Neural entrainment to speech modulates speech intelligibility. *Curr. Biol.* 28, 161–169.
- Tukey, J.W., 1949. Comparing individual means in the analysis of variance. *Biometrics* 5, 99–114.
- Weissbart, H., Kandykaki, K.D., Reichenbach, T., 2019. Cortical tracking of surprisal during continuous speech comprehension. *J. Cogn. Neurosci.* 32, 155.
- Wilsch, A., Neuling, T., Obleser, J., Herrmann, C.S., 2018. Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *Neuroimage* 172, 766–774.
- Zoefel, B., Archer-Boyd, A., Davis, M.H., 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Curr. Biol.* 28, 401.